

Bad Actions or Bad Outcomes? Differentiating Affective Contributions to the Moral Condemnation of Harm

Ryan M. Miller
Brown University

Ivar A. Hannikainen
University of Sheffield

Fiery A. Cushman
Brown University

Moral condemnation of harmful behavior is influenced by both cognitive and affective processes. However, despite much recent research, the proximate source of affect remains unclear. One obvious contender is empathy; simulating the victim's pain could lead one to judge an action as wrong ("outcome aversion"). An alternative, less obvious source is one's own aversion to performing the action itself ("action aversion"). To dissociate these alternatives, we developed a scale that assessed individual aversions to (a) witnessing others experience painful outcomes (e.g., seeing someone fall down stairs); and (b) performing actions that are harmless yet aversive (e.g., stabbing a fellow actor with a fake stage knife). Across 4 experiments, we found that moral condemnation of both first-person and third-party harmful behavior in the context of moral dilemmas is better predicted by one's aversion to action properties than by an affective response to victim suffering. In a fifth experiment, we manipulated both action aversion and the degree of expected suffering across a number of actions and found that both factors make large, independent contributions to moral judgment. Together, these results suggest we may judge others' actions by imagining what it would feel like to *perform* the action rather than *experience* the consequences of the action. Accordingly, they provide a counterpoint to a dominant but largely untested assumption that empathy is the key affective response governing moral judgments of harm.

Keywords: morality, affect, harm, aversion, empathy

Supplemental materials: <http://dx.doi.org/10.1037/a0035361.supp>

In late July 1884, three English sailors stranded in a lifeboat were faced with a grim choice. Having no food, water, or hope of immediate rescue, their best chance at survival was to kill the fourth member of their crew, a severely ill cabin boy, and eat him. The idea seemed unthinkable at first, but the poor conditions of their situation quickly made the threat of death too serious to ignore. Early one morning, while the cabin boy lay unconscious, the captain pulled a penknife from his pocket and sliced through the boy's neck (*R v. Dudley and Stephens*, 1884; Simpson, 1984).

It is hard to read of this grisly act without engaging in moral evaluation. According to recent models, such automatic moral evaluations are strongly influenced, and often determined, by an

emotional response (Greene, 2008; Greene & Haidt, 2002; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Haidt, 2001). Yet, surprisingly little research targets the precise content of this affective response. Some have suggested that it is an empathic reaction to victim suffering (e.g., Crockett, Clark, Hauser, & Robbins, 2010; Pizarro, 2000); thus, in the scenario above, the emotional response would be driven by harm to the boy. More recently, others have proposed that it reflects an aversion to properties of the action itself (Cushman, Gray, Gaffey, & Mendes, 2012; Hannikainen, Miller, & Cushman, 2013); according to this account, the emotional response would be triggered by the act of stabbing rather than concern for the victim. The present study aims to dissociate and evaluate these two sources of affect in the moral domain, thus targeting a matter of central importance in moral psychology: The nature of our affective response to harmful behavior.

The Role of Affect in Moral Judgment

Several lines of convergent evidence from psychology to cognitive neuroscience strongly support a causal role for affect in moral judgment. First, affective reactions to moral violations are often better predictors of condemnation than explicit beliefs about harmful consequences (Haidt & Hersh, 2001; Haidt, Koller, & Dias, 1993). Second, actions widely viewed as impermissible are associated with relatively greater activation in emotion-processing

This article was published Online First February 10, 2014.

Ryan M. Miller, Department of Cognitive, Linguistic, & Psychological Sciences, Brown University; Ivar A. Hannikainen, Department of Philosophy, University of Sheffield, Sheffield; Fiery A. Cushman, Department of Cognitive, Linguistic, & Psychological Sciences, Brown University.

This research was partially supported by a grant from the Office of Naval Research, N00014-13-1-0281.

Correspondence concerning this article should be addressed to Ryan M. Miller, Department of Cognitive, Linguistic, & Psychological Sciences, Brown University, 190 Thayer St., Providence, RI 02906. E-mail: ryan_m_miller@brown.edu

regions of the brain (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene et al., 2001). Third, abnormal patterns of moral condemnation are observed in a variety of clinical populations with known socioemotional deficits, including ventromedial prefrontal cortex lesion patients (Ciaramelli, Muccioli, Lдавас, & Di Pellegrino, 2007; Koenigs et al., 2007), individuals with frontotemporal dementia (Mendez, Anderson, & Shapira, 2005), and psychopaths (Blair, 1995; Koenigs, Kruepke, Zeier, & Newman, 2012). Finally, a growing number of experiments suggest that manipulations of affect alter the perceived permissibility of an action (Crockett et al., 2010; Perkins et al., 2013; Starcke, Ludwig, & Brand, 2012; Starcke, Polzer, Wolf, & Brand, 2011; Valdesolo & DeSteno, 2006; Wheatley & Haidt, 2005; Youssef et al., 2012).

Given the mounting evidence that negative affect contributes to the moral condemnation of harmful actions, we might ask: What is its precise content and origin? The present study takes as its starting point a simple distinction between affective responses grounded in the *outcome* that a behavior causes, versus the *action* that the behavior itself entails.

Our conceptual distinction between action and outcome as sources of affect is closely allied to the theory of dyadic morality (Gray, Waytz, & Young, 2012). According to this theory, prototypical moral situations comprise two roles: an agent and a victim. Agents are associated with the intentional *actions* they perform, and victims are associated with the negative *outcomes* they experience (Gray & Schein, 2012). Evidence of “moral typecasting” further suggests that individuals tend to focus on *either* the mental capacity for agency *or* experience within the same individual, leading to differential effects on judgments of blame (Gray & Wegner, 2009). And, a focus on agents leads to emotions such as anger, whereas a focus on patients leads to emotions such as sympathy (Gray & Wegner, 2011). Collectively, these findings support our proposed distinction between affect associated with an action and affect associated with an outcome. Thus, it is natural to ask which variety of affect is responsible for the aversion to harm evident in the judgment of personal moral dilemmas.

Outcome Aversion

The *outcome aversion* model has substantial support, and it arguably enjoys greater intuitive appeal. Consider again the case of Captain Dudley and the cabin boy. According to the outcome aversion model, a negative moral evaluation of the Captain is grounded in empathy for the boy.¹ You focus on the pain and loss of life that he experiences, and this motivates disapprobation of the captain’s actions. The likelihood that you will judge the action to be wrong overall—that is, even taking into account the utilitarian benefits—would be determined by how personally upset you are by the thought of the boy’s suffering.

Consistent with this model, Hume (1739/2012) argued that moral disapprobation originates in an emotional reaction to the suffering of others, and several modern philosophers and researchers have echoed this view. It is argued that empathy is central to the concept of morality (Blum, 1994; Murdoch, 2001), is critical for moral development (Blair, 1995; Hoffman, 1982, 2001), and plays a crucial role in the process of making moral judgments (Pizarro, 2000).

Further evidence comes from functional neuroimaging. The same regions implicated in processing one’s own pain are also

active during the imagining of another’s pain (Decety, 2011; Singer & Lamm, 2009; Singer et al., 2004), supporting the notion that empathy is directly derived from considerations of the emotional experience of others. Furthermore, the same regions of the “pain matrix” that are activated when one is explicitly instructed to take a victim’s perspective (Decety & Porges, 2011) are activated when one passively views an agent inflicting pain upon a victim (Decety, Michalska, & Akitsuki, 2008), suggesting that one may automatically take the victim’s perspective when encountering third-party transgressions.

Additional evidence consistent with this interpretation comes from the previously discussed effect of serotonin on moral judgment (Crockett et al., 2010): Serotonin only increased condemnation of harm in participants already high in empathy, a result that might be expected if serotonin works by amplifying aversion to the suffering of others.

Action Aversion

The *action aversion* model posits a direct aversion engendered by the action itself. Returning again to Captain Dudley, the model posits that the action of “cutting a throat” is intrinsically aversive, not contingently aversive due to further consideration of the victim’s harm. This focus may be facilitated by a form of perspective taking, or “evaluative simulation,” in which one adopts the agent’s point of view—that is, in which one judges the captain by imagining oneself in a similar circumstance (Hannikainen et al., 2013). In these terms, the negative emotional reaction arises from how bad it would feel to *stab* the boy, rather than how bad it would feel to *be the boy* being stabbed. The action aversion model predicts that the likelihood of judging an action wrong is determined not by empathic concern for the victim, but by how upsetting you consider the action itself to be.

This account also predicts that some actions, like stabbing someone in the neck with a fake knife during a play, might be considered aversive, even though they are ultimately harmless. Consistent with this prediction, Cushman, Gray, Gaffey, and Mendes (2012) found that participants were more physiologically averse to *performing* simulated harmful actions (e.g., shooting an unloaded fake handgun at a consenting experimenter, hitting a rubber hand with a hammer) than they were to *witnessing* a second experimenter perform identical actions, even when they knew that no harm could occur in either case. Levels of aversion were also significantly higher in the *perform* condition than in a *control* condition using metabolically matched actions (e.g., squirting a spray bottle, hitting a wooden block with a hammer). Taken together, these findings demonstrate that “typically violent” actions can possess an aversive quality independent of any imagined harm.

Extending the concept of action aversion to the domain of third-party moral judgment, Hannikainen, Miller, and Cushman

¹ It is worth noting that others (e.g., Batson, Fultz, & Schoenrade, 1987) have drawn a distinction between two senses of “empathy:” an exact mirroring of the affective state of another (e.g., feeling fear when fear is perceived, pain when pain is perceived, etc.), and a response to the affective state of others that is congruent in valence but different in kind (e.g., feeling concerned when fear or pain is perceived). Empathy in either sense could constitute the basis of outcome aversion, because in either case one experiences negative affect linked to the suffering of a victim.

(2013) examined moral judgment directly and found that a tendency to focus on the feelings associated with an action (measured via a self-report scale) predicts the condemnation of harm in personal moral dilemmas, whereas a tendency to focus on outcomes does not. Additional evidence comes from studies targeting moral violations outside of the harm domain. For instance, Lieberman and Lobel (2012) find evidence that one's own aversion to engaging in incest determines one's judgment of third-party incestuous behavior. In fact, the very existence of "victimless crimes" (e.g., incest, cleaning a toilet with a national flag; see Haidt, Bjorklund, & Murphy, 2000; Haidt & Hersh, 2001; Haidt et al., 1993) provides *prima facie* evidence for action aversion; some actions just *feel* wrong, even in the absence of negative consequences.

If it is possible to develop aversions to particular actions, how might this happen? Drawing on observations of animal behavior, Blair (1993, 1995) suggests that normal humans have an innate *violence inhibition mechanism* that initiates a withdrawal response in the presence of victim distress cues. Some aggressive actions reliably cause victim distress, and the mental representations that accompany such actions are thus repeatedly paired with aversive cues. Over time, merely thinking about the action becomes sufficient to trigger an aversive response, and this response is a signal to the agent that the action being considered is *bad*. Thus, although empathy for the victim may be required for the initial development of action aversions, it eventually becomes superfluous to the evaluation of particular actions.

According to this model, action aversion may be acquired for any action property that is typically associated with harm, whether concrete or abstract. At a concrete level, there are certain "typically bad actions," like shooting a gun, thrusting a knife, or shoving a person that are more reliably coupled with victim distress and will therefore become associated with higher levels of aversion. The perceptual and/or motoric properties of such actions will become sufficient to elicit negative affect, even in the absence of harm. As mentioned earlier, this explains why stabbing someone in the neck with a fake knife is considered by many to be aversive. There is also some recent physiological evidence in support of this model. Dillon and Cushman (in prep) asked participants to perform two classes of pretend actions: those which might be thought of as relatively "typical" ways of harming someone in real life (e.g., pointing a gun at someone and pulling the trigger, hitting someone in the foot with a hammer), and more "atypical" forms of these actions that were designed to achieve similar ends (e.g., pulling the trigger of the gun using an attached string, dropping the hammer onto the person's foot using a rope draped over a pulley). Even though the degree of imagined "harm" was equated across conditions, participants showed signs of greater physiological aversion (indicated by a rise in blood pressure) when considering performing the typical actions relative to the atypical actions.

In addition to being associated with low-level, sensorimotor features of an action, aversion might also be triggered at an abstract level by categorical descriptions of actions, such as "murder." As long as the representation, or action category, is associated with negative feedback (e.g., victim distress), it will develop an aversive quality. Although more abstract representations like "murder" are defined partially in terms of their harmful outcomes, this does not necessitate that the affective response they evoke is

due entirely to outcome aversion. A possible source of affect is the negative value attached to the action type, rather than the perceived or imagined suffering of the victim. Experiment 4 explores these ideas in greater detail.

Experimental Approach

The first four experiments in this study are designed to test whether individual differences in action and outcome aversion are predictive of moral judgment in instances of harm. We examine moral judgment in the context of *personal moral dilemmas* (see (Greene et al., 2001) that require participants to judge the permissibility of harming or killing one individual to save the lives of several others (similar to our real-life case of Captain Dudley and his shipmates). We had two primary reasons for using dilemmas with this structure. First, by presenting participants with countervailing considerations, they produce responses with high variability. Second, substantial evidence suggests that affect plays an important role in participants' responses to these dilemmas. Healthy individuals rate them as highly emotionally salient (Koenigs et al., 2007), they activate several regions in the brain associated with emotion (Greene et al., 2001), and manipulations of affect influence perceived permissibility ratings (Starcke et al., 2012; Valdesolo & DeSteno, 2006; Youssef et al., 2012). Thus, they provide ideal testing grounds for measuring the relative importance of our two proposed affective processes to moral judgment. In Experiment 1, we present participants with a series of items designed to assess individual differences in action and outcome aversion and evaluate the contributions of each to moral judgments of harm. In Experiments 2 and 3, we test the robustness of the relationships observed in Experiment 1 by controlling for several potentially confounding factors. In Experiment 4, we explore whether action aversion operates primarily at the sensorimotor level of "typical" violent actions, or whether it can also explain the aversion to more abstract representations of harm.

In addition to observing dispositional differences in action aversion across individuals (as demonstrated in Experiments 1–4), we should also observe systematic differences across *actions*, and these differences should predict the extent to which such actions are morally condemned. Furthermore, this effect should persist when controlling for the magnitude of harm. In Experiment 5, we develop a novel set of stimuli to test this prediction.

Experiment 1

Experiment 1 measures the relative contributions of action aversion and outcome aversion to condemnation of third-party harmful actions in personal moral dilemmas. To assess individual differences in these two types of aversion, we created a number of items to independently target each.

Action/Outcome (A/O) Items

We designed a set of items comprising three types: *action* (9), *outcome* (14), and *control* (11; see Supplementary Materials). The *action* and *outcome* (A/O) items are intended to assess the two aversion types, and the *control* items function as both useful covariates and distractor items (further discussed below). For each item the participant is instructed to indicate how much it would

upset her to perform the *action*, witness the *outcome*, or experience the *control* event described. Thus, we operationalize “aversion” as the extent to which one finds the action, outcome, or control event personally upsetting. Responses are recorded on a 7-point scale from 1 = *Not at all* to 7 = *Very much so*. Three scores are computed for each participant by averaging across the items within each subscale.

One challenge of independently measuring action and outcome aversion lies in the fact that the typical elicitors of each usually co-occur. In other words, when someone performs a violent act, there is usually a suffering victim—the prototypical moral dyad (Gray, Waytz, et al., 2012). Thus, in order to isolate outcome aversion, we asked participants how much it would upset them to simply witness another person experience various types of pain (e.g., “see someone step on broken shards of glass,” or “hear a frightened child crying”). Because these items do not involve a person choosing to perform an action that typically leads to harm, action aversion should not be triggered. Conversely, to isolate action aversion, we stripped typically violent actions of their harmful consequences while preserving essential features of the actions themselves. For example, participants were asked how much it would upset them to “stab a fellow actor in the neck during a play using a stage knife with a retractable blade,” or “shoot a bullet at a consenting friend while he’s behind a bulletproof glass.”² To the extent that such items trigger negative affect in participants, it should be attributable primarily to intrinsic properties of the action itself, rather than anticipated harmful consequences. We note, however, the possibility that some participants may nevertheless be concerned about the potential for such actions to result in harm (e.g., the actor sustaining a neck injury, or their friend being injured by a bullet)—a possibility that may also seem likely given the phenomenon known as *dyadic completion* (Gray, Young, & Waytz, 2012), whereby actions deemed as immoral automatically trigger perceptions of harm, even when harm is absent.

These concerns notwithstanding, there are reasons to believe that our design is capable of isolating the effects of action aversion. First, dyadic completion only applies to actions that are first perceived as *morally wrong*; acting out a scene in a play is not itself immoral, so it is not clear that harm would automatically be inferred. Second, and more importantly, we control for individual differences in outcome aversion in Experiments 1 and 2 (as well as empathic concern in Experiment 2), so any residual relationship between the action items and third-party moral judgment should be attributable to action properties rather than concerns about harm.

In addition to the *action* and *outcome* subscales, we included a *control* subscale. Control items asked participants how upset they would be if particular unfortunate events happened to them, such as “[losing] electrical power for a day,” or “[hearing] the birthday song 100 times in a row.” These items serve two purposes. First, they allow us to statistically control for individual differences in general sensitivity to upsetting events, as well as directional response biases. Second, they draw attention away from the action/outcome items, masking their relation to subsequent moral dilemmas.

Moral Dilemmas

The moral dilemmas employed in this study (see Supplementary Materials) are slightly modified versions of six personal moral dilemmas taken from Greene, Sommerville, Nystrom, Darley, and Cohen (2001) and reworded to reflect a third-party perspective (i.e., targeting what it is permissible for another person to do, rather than what it is permissible for oneself to do). Participants rated the decision to harm one individual in order to save five on a 7-point scale from 1 = *Not morally wrong at all* to 7 = *Very morally wrong*. Ratings were averaged across the six scenarios to compute a single wrongness score for each participant. A larger set of dilemmas including the current six has previously been identified by Koenigs et al. (2007) as “high-conflict,” indicating that they did not elicit 100% agreement from individuals in their study; pretesting of our set of dilemmas using an online sample ($N = 130$) confirmed this classification: mean wrongness ratings for each scenario ranged from 3.75 ($SD = 1.66$) to 4.62 ($SD = 1.74$).

If the condemnation of harm in third-party moral judgment is driven by an aversion to the violent acts themselves, we expect to see a correlation between action item scores and judgments of wrongness for moral dilemmas. Furthermore, this correlation should remain significant after controlling for outcome and control scores if the contribution of action aversion is qualitatively unique. Conversely, if empathy for the victim has a unique influence on the condemnation of harm, we should observe a correlation between outcome item scores and wrongness judgments after controlling for action and control scores. These possibilities are not mutually exclusive, of course, and existing evidence suggests a likely contribution of both sources of affect to the judgment of harmful actions.

Method

Participants. Participants were recruited through both Amazon’s Mechanical Turk (MTurk) and The Moral Sense Test Web site (www.moral.wjh.harvard.edu). Five-hundred and eighty-three participants completed the online survey. Before analysis, participants were excluded according to four criteria: (a) they couldn’t have taken any previous version of the test posted on MTurk, (b) they couldn’t remember having seen any test items at a previous time, (c) they had to spend at least 100 s completing the A/O items, and (d) they had to spend at least 10 s reading and responding to each of the six moral dilemmas. Data meeting these criteria were collected from 330 participants (167 female, three unknown; age group median: 25–34 years).

Procedure. The protocol for this study was approved by the Brown University Institutional Review Board (IRB). All participants voluntarily consented to the study before beginning. They first completed the A/O items followed by six moral dilemmas. The orders of both the A/O items and the moral dilemmas were randomized. Following these two sections, participants were asked whether they had previously encountered any of the test items, and they finished by answering basic demographics questions. All

² We note that the actions described in our action items (e.g., stabbing with a knife, shooting with a gun) may seem more extreme than the events in our outcome items. Pretesting revealed that harm caused by matched actions was judged to be maximally upsetting, so we chose less extreme outcomes that produced greater response variability.

consented participants recruited through MTurk were paid 20–30 cents.

Results

Across participants, the distributions of action ($M = 3.89$, $SD = 1.22$), outcome ($M = 4.77$, $SD = 1.28$), control ($M = 4.44$, $SD = 1.05$), and wrongness ($M = 4.01$, $SD = 1.42$) scores exhibited substantial variance, confirming their suitability for correlational analyses. Each scale also exhibited adequate reliability (Cronbach's alpha: action = .80, outcome = .94, control = .85, moral judgment = .84).³

Both action and outcome scores exhibited significant zero-order correlations with mean wrongness judgments (action: $r(328) = .35$, $p < .001$; outcome: $r(328) = .34$, $p < .001$; Figure 1) although control items did not, $r(328) = .03$, $p = .60$. Notably, all individual action and outcome items were significantly correlated with wrongness judgments, $p < .05$, all $r > .13$. Action and outcome scores were also highly correlated with each other, $r(328) = .64$, $p < .001$.

In a least squares multiple regression model predicting wrongness judgment from action, outcome, and control scores, $R^2 = .16$, $F(3, 326) = 20.21$, $p < .001$, both action score, $\beta = .24$, $t(3,326) = 3.61$, $p < .001$, and outcome score, $\beta = .23$, $t(3,326) = 3.35$, $p = .001$, were uniquely predictive of wrongness.

We further tested a model that included gender, age, and religiosity as covariates, $R^2 = .24$, $F(6, 320) = 16.66$, $p < .001$. Both action score, $\beta = .15$, $t(6,320) = 2.18$, $p < .05$, and outcome score, $\beta = .18$, $t(6,320) = 2.64$, $p < .01$, remained predictive of wrongness after controlling for these other variables.

Discussion

The results of Experiment 1 provide initial support for both the action aversion and outcome aversion hypotheses. Specifically, the tendency to condemn harmful actions in high-conflict personal moral dilemmas was significantly related to (a) how personally aversive one finds “typical” violent but harmless actions, and (b) how upset one becomes when witnessing the pain of others. Furthermore, this relationship persisted when controlling for other variables known to affect moral judgment (gender, age, religiosity). These data are consistent with the idea that when one considers the permissibility of a violent act, she is moved by the emotional aversiveness of both the act itself and its harmful

consequences. Moreover, the results of the multiple regression analysis suggest that the two types of aversion are dissociable and make unique contributions to moral judgment. This is notable considering that at least one model for action aversion considers it to be developmentally dependent on outcome aversion (Blair, 1995).

Although our results are suggestive, one limitation of this experiment arises from the temporal contiguity of the two tasks. Because the A/O items were completed immediately prior to the moral dilemmas, there may have been undue influences of the former on the latter. Experiment 2 uses a lengthy delay period between the two tasks to address this concern.

Experiment 2

The primary purpose of Experiment 2 is to ensure that the relationships between action/outcome scores and wrongness judgments aren't being artificially strengthened by task demands. Because the A/O items and moral dilemmas were presented back to back in Experiment 1, it is possible that responses to the A/O items influenced how participants interpreted and responded to subsequent moral dilemmas. For instance, rating the aversiveness of shooting a gun may have caused the action itself (“gun-shooting”) to become overly salient in moral dilemmas involving guns, and how one responded to the item might have (implicitly or explicitly) informed one's response to the dilemma. Thus, the correlations obtained in Experiment 1 could have been due to the causal influence of A/O items responses on wrongness judgments, rather than being indicative of personal aversions that independently determine responses to both. Furthermore, it is plausible that the very relationship of experimental interest may have been discernible to astute participants.

To eliminate these possible confounds, we administered the A/O items to a group of participants who had judged several moral dilemmas 2–3 years earlier. If moral judgments are normally influenced by action/outcome aversion, and if one's levels of aversion are relatively stable, then we should expect the correlation between A/O items responses and wrongness judgments to persist even across long periods of time. Additionally, we had access to the Empathic Concern subscale of the Interpersonal Reactivity Index (IRI; Davis, 1980, 1983) assessed at the same time as the moral dilemmas. These scores provide a standard measure of participants' dispositional empathy, thus allowing us to further control for other-oriented concern when assessing the relationship of action aversion and moral judgment.

Method

Participants. A subset of participants who had responded to the moral dilemmas and who completed the moral dilemma task (as part of a study conducted 2–3 years earlier) registered with the Moral Sense Test research site by providing their e-mail address, and agreed to be contacted about future studies. To gather data for the A/O items, e-mails containing a link to the questionnaire were sent to all original participants for whom we had contact information.

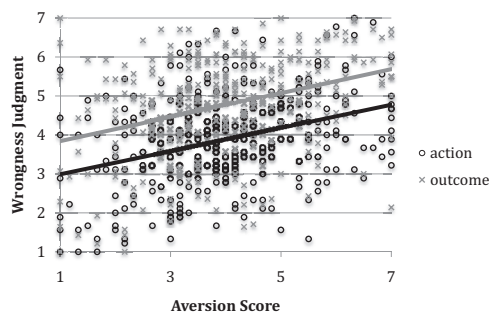


Figure 1. Mean wrongness judgment for personal moral dilemmas by action and outcome scores, Experiment 1.

³ Similar distributions and reliability were obtained in Experiments 2, 3, and 4 and are therefore omitted from further mention.

One-hundred and forty-five participants completed the new A/O survey. Because (a) the study was not conducted on MTurk and (b) we lacked timing data for the moral dilemma questionnaire, the filtering procedure for this experiment was not as stringent as Experiment 1. Our only rule was that participants had to spend at least 100 s completing the A/O items. However, because participants were self-motivated to voluntarily complete both the original study and the A/O items without pay, we expected the quality of responses to be acceptable. Data meeting our criterion were collected from 93 participants (37 females, two unknown; mean age: 36.71 years, $SD = 13.7$).

Procedure. With the exception of A/O item responses, all data for this experiment (moral dilemmas, demographics, IRI) were collected between 2 and 3 years earlier as part of a separate study. The moral dilemmas were a set of 13 “high-conflict” scenarios taken directly from Koenigs et al. (2007) and included slightly reworded versions of the six scenarios used in Experiment 1. However, unlike the scenarios of Experiment 1, all scenarios were written in second person and required the participant to judge the wrongness of committing the harmful action herself. Although it is conceivable that judgments of second- versus third-person behavior might differ, we expect both action and outcome aversion to be relevant to the moral judgment process in either case. Wrongness judgments were made on a 7-pt scale from 1 = *Forbidden* to 7 = *Obligatory*. Because the endpoints of the scale were flipped relative to Experiment 1, all responses were reverse scored to make the results of the two experiments comparable. Relevant items on the IRI were summed to compute scores for the Empathic Concern subscale. Basic demographics were also collected.

Results

We first attempted to replicate our findings in Experiment 1. As Figure 2 shows, wrongness judgments once again correlated with action score, $r(91) = .31, p < .01$. Four out of nine action items were significantly correlated with wrongness ($p < .05$), and all nonsignificant correlations were in the expected direction. However, unlike Experiment 1, outcome score was only marginally related to wrongness judgments, $r(91) = .18, p = .09$, and no individual outcome items reached significance (all $p > .05$). Control score was not related to wrongness, $r(91) = .03, p = .79$. As in Experiment 1, action and outcome scores were strongly correlated, $r(91) = .54, p < .001$. It is also worth noting that outcome scores strongly correlated with empathic concern, $r(90) = .48, p <$

.001, even over a 2–3 year delay, supporting their validity as measures of dispositional aversion to the suffering of others.

The multiple regression model predicting wrongness from action, outcome, and control scores, $R^2 = .10, F(3, 89) = 3.33, p = .02$, further contrasted with Experiment 1. Not only was action score the only significant predictor of moral judgment, $\beta = .32, t(3,89), p = .01$, but the contribution of outcome score was nearly zero, $\beta = .01, t(3,89) = .07, p = .94$ (control: $\beta = -.05, p = .61$).

We also tested a larger model with gender, age, religiosity, and empathic concern (IRI) as additional predictors of wrongness, $R^2 = .17, F(7, 82) = 2.44, p = .03$. Even with this increased complexity, action score remained predictive of wrongness score, $\beta = .31, t(7,82) = 2.26, p < .05$. Notably, empathic concern was also predictive of wrongness, $\beta = .31, t(7,82) = 2.66, p < .05$; however, the empathic concern items and moral judgments were assessed in a concurrent testing session.

Discussion

Despite separating the moral dilemmas and A/O items by over 2 years, action aversion continued to predict the moral judgment of harmful action in the context of dilemmas involving tradeoffs between lives. This finding reinforces the relationship observed between action aversion and moral judgment in Experiment 1, and strongly indicates that is not due to a task-related artifact. Because participants adopted the role of the agent during moral judgment, these data cannot address whether agent perspective-taking is “spontaneous” during judgment of third-party behavior. However, they do provide additional support for the importance of action aversion in moral evaluation.

The robustness of the relationship between action aversion and moral judgment is further evidenced by its survival following inclusion of demographic variables and empathic concern in the regression model. We consider this particularly noteworthy. A few of the action items (e.g., “[hitting] the hand of a corpse with a hammer,” or “[making] obscene gestures at your best friend behind their back”) might be thought of as disrespectful, and one might argue that there is a commensurate degree of “imagined harm” generated by these items. However, if aversion is stemming from the negative consequences, the underlying variance should be shared with outcome aversion and/or empathic concern. The persistent correlation of action aversion and moral judgment after controlling for these factors suggests that the action itself is the source of the aversion.

Experiment 2 failed to replicate the relationship between outcome aversion and wrongness judgments observed in Experiment 1. Their zero-order correlation was marginal, and they shared virtually no unique variance after accounting for action aversion and control scores. What accounts for this discrepancy between Experiments 1 and 2? One possibility is that the correlations observed in Experiment 1 were inflated due to task demands facilitated by the temporal proximity of A/O items and the moral dilemmas. Another possibility is that outcome aversion plays a stronger role in third-party moral condemnation than first-person moral self-regulation. Experiment 3 distinguishes these possibilities while additionally refining several other aspects of our method.

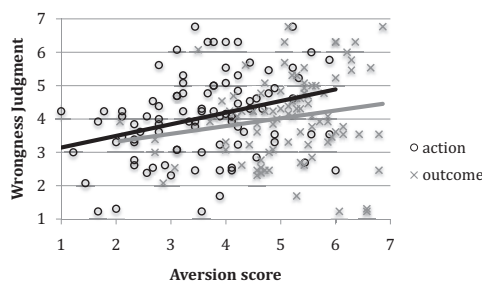


Figure 2. Mean wrongness judgment for personal moral dilemmas by action and outcome scores (collected 2–3 years later), Experiment 2.

Experiment 3

In order to resolve the apparent inconsistency between Experiments 1 and 2, Experiment 3 returns to the use of 3rd party moral evaluation (as in Experiment 1) while attempting to reduce the potential for task demands driving correlation between the A/O items and moral judgment (as in Experiment 2). In order to reduce task demands, we narrowed each of our subscales down to three representative items (discussed below) and introduced 21 filler items plus one “catch” question. By randomly interspersing the three experimental items in each scale with seven times as many filler items, we aimed to mask the relationship between the experimental items and moral dilemmas. The filler items asked participants to rate the intensity of a given emotion elicited by an action or event (e.g., “How upset/happy/scared/angry/excited would you be if. . .”). The affective nature of these questions was intended to provide further camouflage for our experimental items (see Supplementary Materials for the chosen A/O items and accompanying filler items).

Experiment 3 also incorporates two additional refinements to our method. First, while reducing the number of action and outcome items to three each, we took the opportunity to select those items most representative of the constructs of interest. For instance, several of the action items tested in Experiments 1 and 2 (e.g., hitting the hand of a corpse with a hammer, or making rude gestures at your friend behind his back) involve socially objectionable actions that could evoke fears of third-party disapproval if the presence of an observer is imagined. Aversion to such items could also originate from concerns of disrespect, a psychological process potentially more akin to outcome aversion than action aversion. Thus, we chose three action items in which the harmless action was either explicitly consented to by the “receiver,” or consent was implicit in the context (e.g., as part of a play/movie). By making the items less objectionable, we increased the likelihood that any reported aversion would be related to the action itself. The three outcome items were chosen to include both visual and auditory aspects of suffering (e.g., a broken leg, crying). Importantly, there was no advantage of one item type over the other in terms of previous correlations with wrongness judgments across Experiments 1 and 2 (for action items: $r_{mean} = .2$, $r_{min} = .13$, $r_{max} = .26$; for outcome items: $r_{mean} = .21$, $r_{min} = .18$, $r_{max} = .25$).

Second, in Experiments 1 and 2, action and outcome items have been randomly intermixed as part of the same questionnaire. This introduces the possibility that the presence of items in one category influences the judgment of items in the other category, potentially enhancing or suppressing their correlation with moral judgment. This possibility deserves particular attention given that action and outcome scores were strongly correlated across the two previous experiments. Thus, in Experiment 3 we implemented a between-subjects design with action and outcome items presented separately. Following completion of either the action or outcome items, participants were administered the moral dilemma task.

Method

Participants. Participants were recruited online through Amazon’s Mechanical Turk (MTurk). Four-hundred and ninety-five participants completed the online survey. Data were filtered using the same criteria as Experiment 1, with one additional requirement.

Two “catch” items (one inserted among the action/outcome items and one in the demographics section) were included to ensure that participants were paying attention to the task. In order for their data to be used, participants had to answer both items correctly. After filtering data according to the full criteria, we had 73 participants in each of two conditions: action (49 females; age group median = 15–24) and outcome (41 females; age group median: 15–24 years).

Procedure. Following voluntary consent, participants were assigned to either the action or outcome condition. Although assignment to condition was blind to participant details, it was not entirely temporally random; approximately half of the assignments to the outcome condition were made after assignment to the action condition had ended. As in Experiments 1 and 2, all participants responded to six moral dilemmas following completion of the action or outcome items and finished by providing demographic information.

Action and outcome scores were computed for each subject by averaging across the three experimental items within the respective condition. Control scores were computed from only a subset of filler items. We omitted two types of items from this calculation. First, items that ask about the emotional rewards of charitable giving were excluded due to their likely relation to moral judgment. Second, stress and anxiety has been shown to affect moral judgment (Starcke et al., 2012, 2011; Youssef et al., 2012), so all items referencing “fear,” “anxiety,” or “embarrassment” were also excluded. Control scores were therefore based upon 12 filler items (bolded in Supplementary Materials) that bore no *prima facie* relationship with moral judgment.

Results

The zero-order correlations between action/outcome aversion and moral judgment were strikingly different. As Figure 3 demonstrates, action scores were strongly correlated with wrongness score, $r(71) = .49$, $p < .001$, whereas outcome scores were not, $r(71) = .15$, $p = .21$. Moreover, all three action items significantly correlated with wrongness, all $r > .27$, $p < .05$, while no outcome items did, all $r < .20$, $p > .05$.

We next assessed the relationships between action/outcome aversion and wrongness judgments in the presence of additional predictors. Because we employed a between-subjects design, we conducted a separate multiple regression analysis for each condition. In a model predicting wrongness score from action score, control score, gender, age, and religiosity, $R^2 = .39$, $F(5, 66) =$

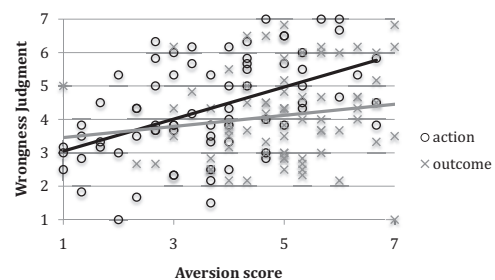


Figure 3. Mean wrongness judgment for personal moral dilemmas by action score and outcome score, Experiment 3.

8.28, $p < .001$, action score remained predictive of wrongness judgment, $\beta = .33$, $t(5,66) = 3.01$, $p = .004$. We evaluated a similar model for participants in the outcome condition, replacing action score with outcome score, $R^2 = .10$, $F(5, 66) = 1.53$, $p = .19$. Consistent with their weak zero-order correlation, outcome score was not predictive of wrongness judgment, $\beta = .05$, $t(5,66) = .35$, $p = .73$.

To determine whether the predictive power of action aversion was significantly stronger than outcome aversion, we also tested for an interaction between condition (action vs. outcome) and aversion ratings. In a multiple regression model predicting wrongness score from *condition* (action vs. outcome), *aversion score*, the *condition* \times *aversion score* interaction, gender, age, and religiosity, the interaction term was significantly predictive, $\beta = .26$, $t(7,136) = 1.99$, $p < .05$, whereas *condition* and *aversion score* were not, both $p > .1$.

Discussion

Combined with the results of Experiment 2, these findings strongly support a role for action aversion in the moral condemnation of harm. It remained predictive of wrongness judgments when measured by a cleaner set of items, camouflaged by superficially similar filler items, and assessed independently of outcome aversion. Furthermore, its predictive power was significantly stronger than that of outcome aversion (as evidenced by the significant interaction term) after controlling for a number of factors with known relationships to moral judgment. As noted previously, the temporal administration of the action and outcome conditions on Mechanical Turk was not perfectly randomized; however, we have no reason to suspect that this introduced a systematic confound into our results.

The results of this experiment, like Experiment 2, call into question the relationship originally predicted between outcome aversion and moral judgment. One potential explanation for this surprising finding is that negative outcomes can occur for *both* the proximal victim *and* the several individuals needing rescue. In other words, one can be averse to the outcome of the harmful action directed toward the single individual, and one can also be averse to the outcome of several individuals dying if no action is taken. Thus, outcome aversions on both sides of the equation might “cancel out” and fail to reliably predict moral judgment in these types of dilemmas. According to this explanation, our failure to find a relationship between outcome aversion and moral judgment is particular to our choice of utilitarian dilemmas as a category of moral judgment.

Still, this null result raises a perplexing question: If empathic concern and outcome items are both measures of dispositional empathy, why is empathic concern correlated with wrongness judgments (as indicated in Experiment 2) yet outcome score is not, or only weakly so (as indicated by Experiments 2 and 3)? We offer two potential explanations for this discrepancy. First, empathic concern was assessed at the same time as the moral dilemmas, so their correlation may be inappropriately high due to the same type of task demand that apparently enhances the ratings of outcome items. A second, distinct possibility is that empathic concern is tapping into more “cognitive” forms of empathy related to specific principles or values (e.g., concerns about social equality and responsibility). For instance, one item on the scale rates how con-

cerned you feel for individuals “less fortunate” than you, and another asks how “protective” you feel toward “someone being taken advantage of.” Outcome aversion and empathic concern (as assessed by the IRI) may therefore be sufficiently dissociable to possess different relationships with moral judgment.

Experiment 4

Our next experiment tests how action aversion applies to increasingly abstract levels of action representation. Experiments 1–3 suggest that it applies at least at a very concrete level: the sensorimotor properties of actions typically associated with harm. Such sensorimotor properties are clearly targeted by the action subscale items, which involve stabbing and shooting at human targets, for instance. Similar sensorimotor properties are also present in most of the dilemmas we used to assess the influence of action aversion on moral judgment; these involved agents who directly, and often violently, inflicted harm upon another individual. It is therefore possible that action aversion is limited to such concrete, sensorimotor representations.

Alternatively, action aversion may also apply to abstract action representations. At a maximally abstract level, for instance, the behaviors described in our moral dilemmas might be categorized as “doing harm.” As discussed in the beginning of the article, such an association is possible if one repeatedly experiences negative feedback (e.g., in the form of victim distress or punishment) while simultaneously holding the abstract representation in mind. Through this process of conditioning, negative affect can become associated with actions that are essentially defined in terms of their goals or foreseen consequences (e.g., “murder,” or “doing harm”) and not simply with specific physical movements or motor plans (e.g., shooting a gun, thrusting a knife).

This analysis requires a conceptual distinction between the aversion to an action construed at the abstract level “doing harm” and an aversion to harmful outcomes. The distinction is subtle but sharp. In the case of outcome aversion, the harmful outcome serves directly as the source of negative affect. That is, the outcome itself is perceived as emotionally aversive, thus motivating nonperformance of the action. In contrast, in the case of action aversion, outcomes may play a role of defining action types, but negative affect is associated with the commission of this abstract action type, not from a value representation associated with the outcome. The contrast might be summarized as the difference between saying “I don’t like harm, so I won’t do it” (outcome aversion) versus “I don’t like doing harm” (action aversion defined over an abstract action type).

There is tentative evidence from our previous experiments consistent with this latter possibility. Two of our six moral dilemmas in Experiments 1 and 3 were scenarios in which “typical” violent actions were not suggested: In one case a person is administered a drug, and in the other he has his kidney removed, but the features of the actions themselves were never specified. As predicted by a broad view in which action aversion is sensitive to more abstract representations, significant correlations between action scores and moral judgment were observed for these two cases.

To test the importance of action aversion in cases of abstract harm, we employed a class of dilemmas known as *impersonal moral dilemmas* (see Greene et al., 2004, 2001). In contrast to personal dilemmas, impersonal dilemmas (a) involve harming an

individual as a side effect rather than a means, and (b) do not require an agent to inflict harm through direct personal force.⁴ Because the actions completely lack aversive surface properties, they can only be construed as harmful at an abstract level (i.e., in terms of their ultimate outcomes).

Why should responses to “impersonal” actions, which lack aversive surface properties, correlate with action subscale items that possess those surface properties but do not involve doing actual harm? The simplest answer is that individuals might differ in their general sensitivity to action-based aversions at any level of abstraction. By analogy, liking apples might correlate with liking oranges because both preferences derive from general individual differences in the enjoyment of fruit.

Let us then take a step back and consider the several sources of variance that might produce correlations between action items and judgment of either personal or impersonal dilemmas. First, as we have seen in our previous three experiments, individuals are differentially averse to actions with specific sensorimotor properties. Because such actions are only present in action items and personal moral dilemmas, this source of variance should only influence correlations between these two item types. Another potential source of variance is the extent to which action aversions influence decision making in general. This differential sensitivity to action values should affect “upsetness” and moral judgments alike, regardless of the level of abstraction. These individual differences should therefore increase correlations between action items and both types of moral dilemmas.

We can now outline three contrasting hypotheses concerning the relationship between action aversion and our two dilemma types. First, we might observe equally strong correlations between action aversion and both personal and impersonal dilemmas. This would indicate that our measure of action aversion is primarily indexing general sensitivities to action values, and the specific sensorimotor properties of actions would be playing little to no role in the relationship between action aversion and moral judgment. Second, we might observe no correlation between action aversion and responses to impersonal dilemmas, suggesting that action aversion is applicable exclusively at the sensorimotor level. Finally, action aversion might be predictive of both personal and impersonal moral judgment, but significantly weaker in the case of impersonal judgment. This observation would be consistent with action aversion operating at both the abstract and sensorimotor level of action representation, with additive effects.

Method

Participants. Four-hundred and fifty-eight participants completed the online survey via Amazon’s Mechanical Turk (MTurk). Data were filtered using the same conservative criteria as Experiment 3. After filtering, there were 65 participants in the *personal* condition (27 females; median age group: 15–24 years) and 96 participants in the *impersonal* condition (46 females; median age group: 25–34 years).⁵

Procedure. After consenting, all participants completed the action items; outcome aversion was not assessed in this experiment. Participants were then administered either six personal moral dilemmas or six impersonal moral dilemmas (see Supplementary Materials). Assignment to the personal/impersonal con-

dition was randomized across both participant and time. Lastly, participants provided demographic information.

Results

Action scores were significantly correlated with wrongness judgments in both the personal, $r(63) = .38$, $p < .01$, and impersonal conditions, $r(94) = .31$, $p < .01$, as shown in Figure 4. Control scores did not correlate with wrongness scores in either condition, $p > .10$.

We further replicated the relationship between action score and personal moral judgment observed in Experiment 3: action score remained predictive of wrongness judgment, $\beta = .31$, $t(5,59) = 2.58$, $p < .05$, when controlling for control score, gender, age, and religiosity. Applying the same model to impersonal moral dilemmas, action score was only marginally predictive of wrongness, $\beta = .22$, $t(5,94) = 1.86$, $p = .06$.

To test whether the relationship with action aversion was significantly weaker in impersonal dilemmas compared to personal dilemmas, we constructed a multiple regression model that included action score, dilemma type, and the action score \times dilemma type interaction, as well as the demographic covariates. We found a significant main effect for dilemma type, $\beta = .52$, $t(7,91) = 6.30$, $p < .001$, but the interaction term was not significant despite a trend in the predicted direction, $\beta = .13$, $t(7,152) = 1.55$, $p = .12$.

Several additional analyses suggest, however, that the true interaction is substantially stronger. First, one highly discrepant observation in the personal moral condition is likely a multivariate outlier, using Mahalanobis distance, $\chi^2(2, N = 65) = 12.36$, $p = .002$. Removing the observation brings the correlation between action score and wrongness judgment closer to that observed in Experiment 3 (updated Experiment 4 $r = .47$; Experiment 3 $r = .49$; see dotted trend line in Figure 4) and results in a significant action score \times dilemma type interaction, $\beta = .19$, $t(7,151) = 2.31$, $p = .02$. Second, if we replace the personal moral condition in Experiment 4 with the data from the action condition of Experiment 3 (an exact replication), the interaction similarly achieves significance, $\beta = .22$, $t(7,159) = 2.84$, $p < .01$. Finally, we can combine data from Experiments 3 and 4 to obtain a better estimate of the relationship between action aversion and personal moral judgment. Even retaining the outlier, there is a significant action score \times dilemma type interaction with this combined dataset, $\beta = .20$, $t(7,224) = 2.53$, $p = .01$.

Discussion

Experiment 4 produced several notable results. First, we replicated the relationship between action aversion and personal moral judgment observed in Experiment 3. Second, we found evidence that action aversion may also be predictive of judgment in impersonal moral dilemmas: action scores and wrongness scores in

⁴ Although Greene et al. (2001) and Greene et al. (2004) do not formulate the personal/impersonal distinction in quite this manner, later work by Greene et al. (2009) targets these properties in particular and demonstrates their impact on moral judgment.

⁵ Despite appearances, the substantial difference in group size (65 vs. 96) was a product of chance and emerged after exclusion criteria were applied.

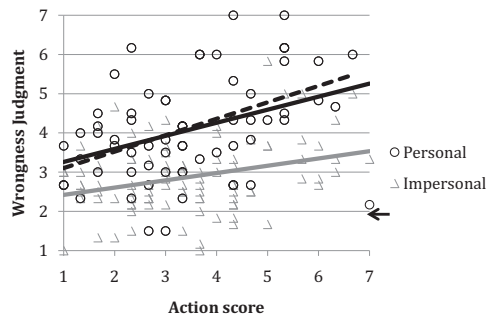


Figure 4. Relationship between mean wrongness judgments and action scores for both personal (black trend line) and impersonal (gray trend line) moral dilemmas. Dashed trend line = personal moral dilemmas when outlier (indicated by arrow) is removed.

impersonal dilemmas were moderately correlated, and action score remained marginally predictive of wrongness after controlling for several demographic variables. Third, we obtained evidence favoring an interaction between action scores and dilemma type; this interaction was significant when (a) a potential outlier was removed, (b) when the personal condition in Experiment 4 was replaced with data from an identical condition in Experiment 3, and (c) when data was aggregated across Experiments 3 and 4, retaining the outlier. Taken together, these findings provide tentative support for the hypothesis that action aversion is relevant at both the sensorimotor and abstract levels of representation.

One potential explanation for the latter finding concerns the sensitivity of action scores to multiple sources of variance. An individual can be more or less averse to specific *types* of actions with typically violent motoric properties, like pushing or shooting, and she can also weight the affective input of an action-based value system more or less heavily when judging the aversiveness of actions. To the extent that such personal aversions influence judgment in personal moral dilemmas, both sources of variance will be important. In contrast, impersonal moral dilemmas lack actions with surface properties that might be considered typically violent, so the relationship between judgments in these dilemmas and action scores will likely be due only to differences in the importance that individuals afford affective signals associated with action-based value representations.

Experiment 5

In this experiment, we adopted a different approach to testing the action aversion hypothesis. Rather than examining how moral condemnation of harm is predicted by *individual differences* in action and outcome aversion (Experiments 1–4), we tested the complementary prediction that moral judgment should be influenced by variation in action aversion brought about through the experimental manipulation of the type of action under consideration. That is, the motoric features of some actions should, on average, be perceived as more aversive than others (consistent with recent physiological evidence; (Dillon & Cushman, in prep)), and this should lead to more severe condemnation of harm brought about by these actions. Furthermore, this effect should be independent of outcome-related concerns, such as beliefs about the degree of suffering or harm endured by the victim.

The harmful actions in this experiment were performed in the context of consensual, hypothetical “mercy killings.” Participants were told to imagine a variety of ways that a terminally ill man might request being killed (e.g., via pill administration, suffocation by pillow, being shot in the head, etc.; see Supplementary Materials for the complete list of stimuli). One group of participants rated how *morally wrong* it would be for a third party to agree to kill the man in each way, and the average wrongness rating for each killing method constituted the dependent variable. A second group imagined carrying out the mercy killings themselves and rated how much *suffering* they thought each method would cause. Because participants were specifically told that each method would ultimately kill the man, these suffering ratings were intended to capture the variance between killing methods associated with outcome-related concerns. A final group was told to imagine that they were actors in a movie, and that the mercy killings were therefore entirely scripted and fake. Furthermore, they were told that absolutely *no harm* would come to the actor playing the victim. This group rated how much it would upset them to act out fake versions of each mercy killing, thus targeting the level of “action aversion” associated with each method.

This between-subjects experimental design offered several methodological advantages. First, the items used to assess action and outcome aversion were identical to those for which wrongness judgments were obtained. This allowed us to determine the precise contributions of action and outcome aversion to the moral judgment of a variety of action types. Second, it involves an experimental manipulation of action aversion across stimuli, rather than a correlational finding driven by individual differences in sensitivity to action aversion. Meanwhile, our manipulation of action and outcome properties did not depend on a priori assumptions about which items ought to be judged high or low in aversiveness. Rather, participants provided these ratings for us; we designed the stimuli merely with an eye toward variety. Third, the mercy killing context allowed for a cleaner test of the outcome aversion hypothesis. Because harm was limited to a single individual (unlike in our moral dilemmas, where harms toward many people were traded off against each other), a source of concern that qualified our interpretations of Experiments 1–4 was avoided.

Method

Participants. Participants were recruited online through Amazon’s Mechanical Turk (MTurk). One-hundred and ninety-five participants completed the online survey. Data from these participants were filtered according to two criteria. After viewing instructions, all participants were given a multiple-choice question testing their comprehension of the task; only those participants who answered correctly were included in the analyses. Participants were also required to spend at least 60 s responding to the 23-item questionnaire in each condition.

After filtering, there were 48 participants in the action condition (25 female; mean age: 33.46 years), 55 participants in the outcome condition (25 female, one unknown; mean age: 35.07 years), and 50 participants in the moral judgment condition (31 female; mean age: 36.72 years).

Procedure. After consenting, each participant was randomly assigned to one of three conditions: *moral judgment*, *outcome*, or

action. Each condition was associated with judgments of 23 items (described below), followed by basic demographic questions.

Those assigned to the *moral judgment* condition were told to imagine a terminally ill man named John who sincerely wanted to die. Furthermore, they were told that John had asked a second individual, Carl, to perform a mercy killing, and that Carl had taken pity on John and agreed to carry out his wish. Participants were then presented with 23 ways that John might request to be killed, and they were told to assume that John would, in fact, die. For each method, participants rated *how morally wrong* it would be for Carl to kill John in that particular manner. Ratings were provided on a Likert scale ranging from 1 (*the LEAST wrong it could be to kill someone*) to 10 (*the MOST wrong it could be to kill someone*).

The instructions for the *outcome* condition were similar to the moral judgment condition, with two important differences. Instead of being told that a third party would perform the mercy killing, participants were told to imagine killing John themselves; this was done to parallel the first-person perspective used in the *action* condition. Participants were then asked to rate how much suffering each of the 23 methods would be expected to cause on a scale from 1 (*the LEAST suffering John could experience*) to 10 (*the MOST suffering John could experience*).

In contrast to the previous two conditions, participants in the *action* condition were told to imagine carrying out a *fake* mercy killing as part of a movie plot. The instructions emphasized both the “pretend” nature of the action, and the fact that no harm would come to the actor playing John—that is, the movie set and props had been constructed in such a way that harm was impossible. Participants rated how much it would upset them to act out each of the 23 variants of mercy killing on a scale ranging from 1 (*the LEAST it could upset me to pretend kill someone*) to 10 (*the MOST it could upset me to pretend kill someone*).

Wrongness, outcome, and action scores were computed for each method of mercy killing by averaging across participants’ ratings within the respective condition. This resulted in three vectors of 23 values (one value for each action type) to be used in subsequent correlational and regression analyses.

Results

Averaging across participant responses within each condition, the item-by-item action scores (min = 3, max = 7.44, $M = 5.71$, $SD = 1.15$), outcome scores (min = 3.45, max = 9.44, $M = 7.26$, $SD = 1.51$), and wrongness scores (min = 3.6, max = 9.48, $M = 7.12$, $SD = 1.45$) all exhibited substantial variance, indicating that action aversion, perceived degree of harm, and moral wrongness had been successfully manipulated across the 23 methods of mercy killing.

Notably, action scores and outcome scores were not significantly correlated with each other, $r(21) = .13$, $p = .57$. However, both action and outcome scores exhibited very strong (and nearly equal) zero-order correlations with wrongness judgments (action: $r(21) = .66$, $p < .001$; outcome: $r(21) = .68$, $p < .001$; Figure 5). A least squares multiple regression model revealed that, together, action and outcome scores explained nearly 80% of the variance in wrongness judgments, $R^2 = .79$, $F(2, 20) = 36.95$, $p < .001$. As would be expected from their nonsignificant correlation, the contributions from both action score, $\beta = .58$, $t(2,20) = 5.58$, $p <$

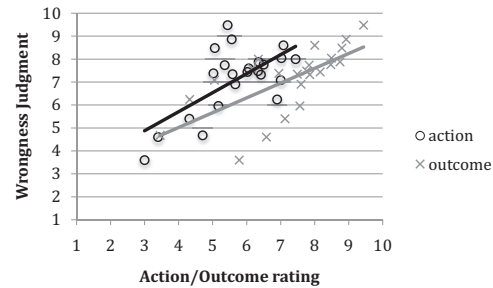


Figure 5. Moral wrongness predicted by action aversion (action) and beliefs about suffering (outcome). Note: Unlike previous figures that plotted individual subject scores, each datapoint here represents the average score for a particular action type.

.001, and outcome score, $\beta = .60$, $t(2,20) = 5.78$, $p < .001$, were largely unique.

Discussion

The results of Experiment 5 showed that in the context of third-party mercy killing, moral condemnation of harmful action was almost entirely explained by two independent factors: the degree of suffering the action was expected to cause, and the relative aversiveness of the action itself (even in the absence of harm). In contrast to Experiments 2 and 3, these findings provide support for the importance of both action and outcome aversion in the moral judgment of harm, while also demonstrating their dissociable influences.

These data strongly suggest that action aversion cannot be reduced to concerns about harm. Participants in the action condition were explicitly told that no harm would come to the actor playing John; yet, they still found certain actions to be substantially more upsetting than others. Action aversion ratings were also not correlated with the degree of suffering that each mercy killing method was believed to cause, and they continued to predict moral judgment when controlling for perceived degrees of suffering. Even if aversion in the action condition reflected some other type of outcome-oriented concern, such as the inherent *danger* of each action (i.e., the probability that the action *could* cause harm), such variance would not be applicable to the moral judgment condition; participants who provided moral judgments were expressly told that each action would in fact kill John, eliminating any ambiguity concerning the “effectiveness” of each method. Thus, the strong relationship observed between the action condition and moral judgment condition is best explained in terms of the surface properties of the actions that are common to each.

Equally notable was the strong correlation between suffering ratings and moral judgment. Whereas our earlier attempts failed to detect a consistent effect at the level of individual differences, these results confirm that outcome considerations can have a substantial impact on moral judgment. One potential explanation for the apparent discrepancy between experiments concerns structural differences in the stimuli. Because the harm associated with mercy killings is directed at a single individual, being averse to harm may compel condemnation of the killing in a way that it does not in personal moral dilemmas, where harm will inevitably result either to the focal individual or else to a group of several others. A

second possibility depends on the different outcome measures used in each experiment. In Experiments 1–3, outcome aversion was determined by asking participants how *upsetting* various harmful outcomes were, whereas the current experiment asked participants about the *degree* of suffering that would result from each action. Although affective responses to harm should track beliefs about the magnitude of harm, the latter may also be a source of purely cognitive, or informational, influences on moral judgment.

The data from Experiment 5 also allow us to rule out the possibility that variance in moral judgment determines the aversion to performing an action, rather than the other way around—that is, that people are averse to pretending to stab in a play because of their judgment that it is morally wrong to stab in real life. We found that ratings of victim suffering were a strong predictor of moral judgments of mercy killing, but did not significantly predict the aversion to properties of the action performed as theater, as would have been expected if moral judgment determined action aversion rather than the converse.

General Discussion

The present study tests two potential source of affect in moral judgments of harm. The *outcome aversion* hypothesis asserts that the condemnation of harmful actions derives from outcome-oriented concerns, such as an empathic response to the victim, whereas the *action aversion* hypothesis states that condemnation is driven by an aversive response to the action itself. Across five experiments, we found consistent and strong support for the importance of action aversion in the context of moral dilemmas. The extent to which individuals found the performance of harmless actions with characteristically violent features upsetting was a consistent predictor of their willingness to condemn harmful, utilitarian actions. This relationship was also markedly robust: It endured over a period of more than two years (Experiment 2), and it was observed when employing both a large number of experimental action items (Experiments 1 & 2) and a smaller number of more highly controlled items (Experiments 3 & 4). We also found evidence that action aversion varies substantially across action types, and the average level of aversion associated with an action strongly predicts how wrong that action is judged to be (Experiment 5). Furthermore, the effect persists when controlling for beliefs about the magnitude of suffering caused by each action.

These findings join a cluster of recent others (Cushman et al., 2012; Hannikainen et al., 2013; Lieberman & Lobel, 2012) in shedding new light on the psychological processes that support moral condemnation. When judging the wrongness of another individual's harmful action, we appear to engage in a process of "evaluative simulation," whereby we evaluate the moral status of another's action by simulating the affective response that we would experience performing the action ourselves. In this way, first-person aversions normally involved in self-regulation can directly influence third-party moral condemnation.

We have interpreted our results to favor a unidirectional causal effect of action and outcome aversion on moral judgment. Although it is possible that a third variable independently determines both personal aversions and moral intuitions (with no direct link between the two), we believe the causal interpretation is most consistent with the evidence for evaluative simulation. Hannikainen et al. (2013) found that the relationship between first-

person disgust sensitivity and condemnation of third-party purity violations was moderated by perspective taking: Individuals who reported taking the agent's perspective during moral judgment exhibited a stronger relationship between their own aversions and third-party condemnation. Such a result is to be expected if action aversions directly inform moral judgment, but is not readily explained by a "third variable" theory. Similarly, Lieberman and Lobel (2012) employed mediation analyses to demonstrate that first-person aversions to incestuous behavior causally influence condemnation of third-party incestuous acts. Although distinct considerations likely underlie transgressions involving harm and deviant sexual behavior or purity concerns, these observations provide direct evidence that personal aversions can influence moral judgment of others' behavior. It has even been noted that such influence may confer a strategic advantage; after all, there are few drawbacks to endorsing and spreading norms that prohibit actions one has no desire to perform anyway (Tybur, Lieberman, Kurzban, & DeScioli, 2013). Combining these results with ours, we suggest that action aversion may be integral to the moral judgment process across a variety of contexts and moral concerns. Indeed, Hannikainen et al. (2013) found that increased tendency to focus on actions during moral judgment was associated with greater perceived importance of each of the five moral foundations proposed by Haidt and Graham (2007).

In addition to explaining individual differences in the judgment of moral dilemmas, action aversion provides a natural explanation for two peculiar, nonutilitarian features of moral judgment: All else being equal, harm caused by direct personal force is judged worse than harm mediated by an external force (e.g., a gun; Greene et al., 2009; also similar to the *contact principle* in Cushman et al., 2006), and harmful actions are judged to be worse than harmful omissions (Baron & Ritov, 2004; Cushman et al., 2006; Spranca, Minsk, & Baron, 1991). The *personal/external force* distinction is a simple by-product of the way that action aversions are formed: The strongest aversions will be associated with actions that are repeatedly and reliably paired with victim distress, and many of these actions—for example, pushing, punching, kicking, slapping, stabbing—happen to involve the direct transfer of personal force. The *action/omission* distinction falls out of the fact that actions have properties that can be associated with harm, whereas omissions lack such specific properties. The choice to "do nothing" is not generally associated with either good or bad outcomes, in much the same way that mundane activities like flipping a switch are not. Thus, the psychological mechanism that forms action aversions only associates negative value with actions, and not the omission of action. Without any affective tags, omissions do not possess the same aversive quality as actions and are therefore perceived as less wrong.

Our claim that first-person aversions exert an influence during the evaluation of third-party dilemmas makes a key prediction that we do not test here: That affective systems are engaged during the process of third-party moral judgment. We rely on the observation from previous studies (e.g., Koenigs et al., 2007; Tassy et al., 2012) that participants (a) report experiencing emotion during similar moral dilemmas, and (b) increasingly condemn actions in dilemmas rated high in emotional intensity. However, future studies employing our method might assess whether individual differences in the subjective emotional intensity of moral dilemmas are

better predicted by action scores than outcome scores, as our hypothesis would predict.⁶

As we noted above, it is natural to associate the distinction between action- and outcome-based affect targeted in the present study with the distinction between agent and patient roles targeted in the dyadic theory of morality (Gray, Waytz, et al., 2012). However, it is also worth noting the distinctions between this previous work and the current approach. The theory of dyadic morality focuses primarily on cognitive processes of mind perception, describing how moral judgments arise from ascriptions of *agency* or *experience* to members of the moral dyad. By contrast, the current approach focuses on decision-making processes that derive affect either from a contemplated action or its expected outcome. These decision-making processes are not restricted to the moral domain. For instance, a person presented with fudge shaped like feces (Rozin, Millman, & Nemeroff, 1986) might feel torn between aversion to an action (“I can’t bring myself to put this in my mouth . . .”) and knowledge of its outcome (“ . . . although I know that it will taste delicious.”) Yet, framed in this context, the rival decision-making systems obviously do not require mind perception.

Thus, it remains a critical area for future research to explore the connections between two sources of affect (action vs. outcome), two dimensions of mind (agent and patient), and two roles in the moral dyad (perpetrator and victim). One question of particular importance is whether an agent must perceive his *own* mind as agentic in order to be averse to performing harmful action. A second is whether action-based value representations in the moral domain necessarily entail the perception or inference of an *other’s* suffering mind (Gray, Young, et al., 2012), or whether they may sometimes be linked to sensorimotor properties of the action in isolation (Cushman et al., 2012).

This latter question also highlights a central issue to be addressed in the study of action aversion: the precise mechanism by which action representations elicit an aversive response. We have proposed that negative affect, through a process of conditioning, becomes directly associated with mental representations of actions that routinely cause harm. An action representation therefore becomes a type of “conditioned stimulus,” capable of independently eliciting the aversion originally triggered by distress cues, punishment, or other types of negative stimuli that were paired with the action in the past (Cushman, in press). Yet, researchers have long known that Pavlovian conditioning involves more than the shifting of a behavioral response from one stimulus to another (Rescorla, 1988); it can also encode the complex *relations* that hold among various stimuli. Thus, the mental representation of an action might automatically trigger an “implicit” representation of the harm commonly associated with the action, and this implicit representation might itself be a source of negative affect. Further research will be required to distinguish between these two accounts, and to determine whether other types of Pavlovian associations might also impinge on moral judgment (Crockett, 2013).

Although our principle interest is in the role of action-based affective processes in moral judgment, the apparent lack of a role for outcome-based affective processes in Experiments 2 and 3 is so striking that it demands attention. Given that the capacity for empathy is held by many to be a critical component of moral behavior and judgment (e.g., Hoffman, 1982, 2001; Pizarro, 2000), why did we find limited evidence for this relationship? One might

argue that this result is even more paradoxical given our own favored explanation for the development of action aversion; if sensitivity to the distress cues of others (i.e., outcome aversion) is necessary for action aversions to form (Blair, 1993, 1995), how could this sensitivity not also be present during moral judgment?

As noted earlier, this finding may derive from our methodological choice of dilemmas that present a utilitarian rationale for harmful action. Because one is not required to be the proximate cause of negative outcomes in order for them to be considered aversive, outcome aversion may equally favor action (to avoid the suffering of many) and inaction (to avoid the suffering of one). It may seem, *prima facie*, that outcome aversion would steer one toward the utilitarian choice because the death of many should be more emotionally aversive than the death of one. Several studies suggest, however, that people find the suffering of a group of individuals to be no more—and perhaps even *less*—emotionally distressing than the suffering of a single individual (Kogut & Ritov, 2005a, 2005b; Slovic, 2010). Furthermore, empathy may be moderated by physical distance (Latané, Liu, Nowak, Bonevento, & Zheng, 1995; Latané, 1981; Milgram, 1965), making the suffering of the proximal victim relatively more salient. For these reasons outcome aversion may not consistently favor one option over the other when faced with opposing negative consequences. The findings of Experiment 5 support this interpretation: When harm was limited to a single individual (in the context of mercy killings), the amount of suffering associated with an action *was* a significant predictor of its moral wrongness.

A second possibility is that our measure of outcome aversion is not sufficiently related to the empathic response normally elicited by violent actions. For instance, most of the outcome items focused on accidental/unintentional harms, and there is evidence that unintentional harms are perceived as less severe than intentional harms (Gray & Wegner, 2008). Our rationale for choosing unintentional harms was to pinpoint affective responses to victim suffering untainted by consideration of any potentially aversive action (a necessary component of intentional harm). And, in any event, we found that outcome aversion strongly correlated with the empathic concern subscale of the IRI after 2 years, increasing our confidence in its validity as a measure of affective empathy.

A final possibility is that outcome aversion is a normal, automatic response to the suffering of others, but that *moral judgment*, in particular, refocuses attention on the aversiveness of the action rather than the suffering of the victim. This account would reconcile the current results with neuroimaging evidence that suggests witnesses of violence spontaneously adopt the victim’s perspective (Decety et al., 2008; Decety & Porges, 2011). Making a moral judgment might promote *agent* perspective taking and/or action focus in a way that passively viewing victims befalling harm does not. Thus, outcome aversion might affect moral judgment only indirectly, playing a critical role in the acquisition of action aversions, but exerting little influence on the judgment process itself. In other words, empathy could be important for moral develop-

⁶ A second limitation is our focus on a single dependent variable, namely, moral wrongness. Other types of judgments in which outcome considerations play a larger role, such as moral blame or punishment, might bear different relationships with our action/outcome measures; testing these independently will provide a clearer understanding of the role that action aversion plays in moral cognition more broadly.

ment without being necessary for moral judgment (a point also made by Prinz, 2011).

Note that this account does not preclude a role for outcome considerations in the process of moral judgment. To the contrary, the expected outcomes of an action clearly matter: Stabbing a fellow actor with a fake, harmless knife is perfectly permissible, whereas using a real knife would be severely condemned. It may be that beliefs about harm primarily serve as “cognitive” input to the moral judgment process, rather than exerting their influence via emotional/affective pathways. Interestingly, this possibility is compatible with the relationship observed between expectations of suffering and judgments of wrongness in Experiment 5. In contrast to Experiments 1–4, which targeted emotional aversion specifically, Experiment 5 assessed *beliefs* about the magnitude of suffering. Although we expect this measure to capture variance related to the emotional aversiveness of harm (i.e., the more a victim suffers, the more aversive it should be, on average, to an observer), it might also reflect more “cognitive” representations that could inform moral judgment even in the absence of felt aversion. Future studies will be necessary to choose between these explanations, among others.

Conclusion

The present study provides a novel approach to assessing affective contributions to the moral condemnation of harm. We found that moral evaluations are more heavily influenced by an aversion to the act of harming than by an emotional reaction to harm itself. Consequently, associating negative value with harmful actions may play a key role in developing normal moral intuitions, a point that may ultimately have important implications for clinical populations, such as psychopaths, who exhibit deficits in both aversive conditioning (Blair, 2001) and moral judgment (Blair, 1995).

References

- Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, 94, 74–85. doi:10.1016/j.obhdp.2004.03.003
- Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, 55, 19–39. doi:10.1111/j.1467-6494.1987.tb00426.x
- Blair, R. J. R. (1993). *The development of morality* (Unpublished doctoral dissertation). University of London, London, UK.
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57, 1–29. doi:10.1016/0010-0277(95)00676-P
- Blair, R. J. R. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry*, 71, 727–731. doi:10.1136/jnnp.71.6.727
- Blum, L. A. (1994). *Moral perception and particularity*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511624605
- Ciamarelli, E., Muccioli, M., Lådavas, E., & Di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2, 84–92. doi:10.1093/scan/nsm001
- Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, 17, 363–366. doi:10.1016/j.tics.2013.06.005
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17433–17438. doi:10.1073/pnas.1009396107
- Cushman, F. (in press). Action, outcome and value: A dual-system framework for morality. *Personality and Social Psychology Review*.
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, 12, 2–7. doi:10.1037/a0025071
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment. *Psychological Science*, 17, 1082–1089. doi:10.1111/j.1467-9280.2006.01834.x
- Davis, M. H. (1980). A multi-dimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113–126. doi:10.1037/0022-3514.44.1.113
- Decety, J. (2011). Dissecting the neural mechanisms mediating empathy. *Emotion Review*, 3, 92–108. doi:10.1177/1754073910374662
- Decety, J., Michalska, K. J., & Akitsuki, Y. (2008). Who caused the pain? An fMRI investigation of empathy and intentionality in children. *Neuropsychologia*, 46, 2607–2614. doi:10.1016/j.neuropsychologia.2008.05.026
- Decety, J., & Porges, E. C. (2011). Imagining being the agent of actions that carry different moral consequences: An fMRI study. *Neuropsychologia*, 49, 2994–3001. doi:10.1016/j.neuropsychologia.2011.06.024
- Dillon, K., & Cushman, F. (in prep). *Typical vs. atypical harm*. Manuscript in preparation.
- Gray, K., & Schein, C. (2012). Two minds vs. two philosophies: Mind perception defines morality and dissolves the debate between deontology and utilitarianism. *Review of Philosophy and Psychology*, 3, 405–423. doi:10.1007/s13164-012-0112-5
- Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, 23, 206–215. doi:10.1080/1047840X.2012.686247
- Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science*, 19, 1260–1262. doi:10.1111/j.1467-9280.2008.02208.x
- Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, 96, 505–520. doi:10.1037/a0013748
- Gray, K., & Wegner, D. M. (2011). Dimensions of moral emotions. *Emotion Review*, 3, 258–260.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101–124. doi:10.1080/1047840X.2012.651387
- Greene, J. D. (2008). The secret joke of Kant’s soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol 3: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–80). Cambridge, MA: MIT Press.
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111, 364–371. doi:10.1016/j.cognition.2009.02.001
- Greene, J. D., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6, 517–523. doi:10.1016/S1364-6613(02)02011-9
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400. doi:10.1016/j.neuron.2004.09.027
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108. doi:10.1126/science.1062872

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834. doi:10.1037/0033-295X.108.4.814
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason*. Unpublished manuscript.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98–116. doi:10.1007/s11211-007-0034-z
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31, 191–221. doi:10.1111/j.1559-1816.2001.tb02489.x
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628. doi:10.1037/0022-3514.65.4.613
- Hannikainen, I., Miller, R., & Cushman, F. (2013). *Act versus impact: Conservatives and liberals exhibit different structural emphases in moral judgment*. Manuscript submitted for publication.
- Hoffman, M. L. (1982). Development of prosocial motivation: Empathy and guilt. In N. Eisenberg (Ed.), *The development of prosocial behavior* (pp. 281–313). New York, NY: Academic Press.
- Hoffman, M. L. (2001). *Empathy and moral development: Implications for caring and justice*. Cambridge, UK: Cambridge University Press.
- Hume, D. (2012). *A treatise of human nature*. Mineola, NY: Courier Dover Publications. (Original work published 1739)
- Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, 7, 708–714. doi:10.1093/scan/nsr048
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. R. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446, 908–911. doi:10.1038/nature05631
- Kogut, T., & Ritov, I. (2005a). The “identified victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, 18, 157–167. doi:10.1002/bdm.492
- Kogut, T., & Ritov, I. (2005b). The singularity effect of identified victims in separate and joint evaluations. *Organizational Behavior and Human Decision Processes*, 97, 106–116. doi:10.1016/j.obhdp.2005.02.003
- Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36, 343–356. doi:10.1037/0003-066X.36.4.343
- Latané, B., Liu, J. H., Nowak, A., Bonevento, M., & Zheng, L. (1995). Distance matters: Physical space and social impact. *Personality and Social Psychology Bulletin*, 21, 795–805. doi:10.1177/0146167295218002
- Lieberman, D., & Lobel, T. (2012). Kinship on the Kibbutz: Coresidence duration predicts altruism, personal sexual aversions and moral attitudes among communally reared peers. *Evolution and Human Behavior*, 33, 26–34. doi:10.1016/j.evolhumbehav.2011.05.002
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology*, 18, 193–197. doi:10.1097/01.wnn.0000191292.17964.bb
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, 18, 57–76. doi:10.1177/001872676501800105
- Murdoch, I. (2001). *The sovereignty of good*. 1970. London, UK: Routledge.
- Perkins, A. M., Leonard, A. M., Weaver, K., Dalton, J. A., Mehta, M. A., Kumari, V., . . . Ettinger, U. (2013). A dose of ruthlessness: Interpersonal moral judgment is hardened by the anti-anxiety drug lorazepam. *Journal of Experimental Psychology: General*, 142, 612–620. doi:10.1037/a0030256
- Pizarro, D. (2000). Nothing more than feelings? The role of emotions in moral judgment. *Journal for the Theory of Social Behaviour*, 30, 355–375. doi:10.1111/1468-5914.00135
- Prinz, J. (2011). Is empathy necessary for morality? In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and psychological perspectives* (pp. 211–229). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199539956.003.0014
- Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43, 151–160. doi:10.1037/0003-066X.43.3.151
- Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, 50, 703–712. doi:10.1037/0022-3514.50.4.703
- R v. Dudley and Stephens., 14 QBD 273 (1884).
- Simpson, A. W. B. (1984). *Cannibalism and the common law* □: *The story of the tragic last voyage of the Mignonette and the strange legal proceedings to which it gave rise*. Chicago, IL: University of Chicago Press.
- Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences*, 1156, 81–96. doi:10.1111/j.1749-6632.2009.04418.x
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303, 1157–1162. doi:10.1126/science.1093535
- Slovic, P. (2010). If I look at the mass I will never act: Psychic numbing and genocide. *Emotions and Risky Technologies*, 5, 37–59. doi:10.1007/978-90-481-8647-1_3
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27, 76–105. doi:10.1016/0022-1031(91)90011-T
- Starcke, K., Ludwig, A.-C., & Brand, M. (2012). Anticipatory stress interferes with utilitarian moral judgment. *Judgment and Decision Making*, 7, 61–68.
- Starcke, K., Polzer, C., Wolf, O. T., & Brand, M. (2011). Does stress alter everyday moral decision-making? *Psychoneuroendocrinology*, 36, 210–219. doi:10.1016/j.psychen.2010.07.010
- Tassy, S., Oullier, O., Duclos, Y., Coulon, O., Mancini, J., Deruelle, C., . . . Wicker, B. (2012). Disrupting the right prefrontal cortex alters moral judgement. *Social Cognitive and Affective Neuroscience*, 7, 282–288. doi:10.1093/scan/nsr008
- Tybur, J. M., Lieberman, D., Kurzban, R., & DeScioli, P. (2013). Disgust: Evolved function and structure. *Psychological Review*, 120, 65–84. doi:10.1037/a0030778
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17, 476–477. doi:10.1111/j.1467-9280.2006.01731.x
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16, 780–784. doi:10.1111/j.1467-9280.2005.01614.x
- Youssef, F. F., Dookeeram, K., Basdeo, V., Francis, E., Doman, M., Mamed, D., . . . Legall, G. (2012). Stress alters personal moral decision making. *Psychoneuroendocrinology*, 37, 491–498. doi:10.1016/j.psychen.2011.07.017

Received February 26, 2013

Revision received October 2, 2013

Accepted October 3, 2013 ■